USING MULTIVARIATE DATA ANALYSIS FOR PROCESS TROUBLE SHOOTING

Patricia Winchell

Catalyst Paper Corporation
Crofton Division
PO Box 70
Crofton BC V0R 1R0

ABSTRACT

Multivariate data analysis tools were used to improve the understanding of the wet end chemistry and white water system of the Papermill at NorskeCanada Crofton Division. Specifically, the analysis was aimed at identifying what variables were contributing to increased retention aid use and wet end instability. Several models were developed using data sets with up to 88 process variables and over 3000 observations. It was found that increased retention aid use was driven primarily by PCC and TMP usage as well as the addition of Alaskan White Spruce to the TMP furnish.

INTRODUCTION

Multivariate Data Analysis (MVDA) is a powerful tool that can be used to trouble shoot complicated operational problems in the pulp and paper industry.

In the work discussed in this paper, multivariate analysis tools such as principle components analysis (PCA) and partial least squares projections to latent structures (PLS) were used to improve the understanding of increased retention aid chemical use on NorskeCanada Crofton Division's #3 Paper Machine.

BACKGROUND

NorskeCanada Crofton Division began as a single-line kraft pulp mill in 1957. Today Crofton's three paper machines and two pulp machines have an annual capacity of 680,000 tonnes of product each year. Crofton is an industry leader in the production of high-quality lightweight papers and produces both newsprint and lightweight telephone directory paper for customers all over North America, Western Europe, Asia and Latin America.

The principle furnish used on the paper machines are hydrosulphite bleached TMP, De-Inked Pulp, Kraft and Precipitated Calcium Carbonate. The mill has successfully used a cat-cat retention aid system for the past number of years. The flocculant is added to the thin stock just after the primary screen and also to the save-all. Coagulant is added to the suction side of the primary fan pump and also by a ratio to the main flocculant addition point. An online white water solids analyzer controls the main flocculant addition rate. The main coagulant addition is controlled by the stock preparation operators based on regular cationic demand tests.

METHODS AND MATERIALS

The initial step in performing a multivariate analysis is to determine what the overlying question is that you would like to ask of the data. The scope of this question will then determine the variety and amount of data that is required for the analysis. For the discussion in this paper the question that is focused on is 'What is contributing to increased retention aid use and chemical instability on #3 Paper Machine'.

To answer this question, a variable list was developed incorporating the most relevant process variables including stock flows & consistencies, chemical addition rates, white water cationic demand measurements, couch vacuum levels, chest level outputs, machine speeds and paper grades. All of these variables are continuously monitored by the mills digital control system and archived in the mills data historian. A spreadsheet was populated with the data at regular observation times from the data historian. Care was taken to ensure that the sample observation times were far enough apart to incorporate any transport time between the process units.

The most crucial part of performing a good multivariate data analysis is the initial cleaning and organizing of the data. First, the completed spreadsheet was thoroughly reviewed to ensure that all of the results were reasonable and that there were no problems with the extraction of the sample data from the data historian. Second, to simplify the future analysis, each variable was given a short, distinct name. For example, an instrument tag name such as 'FC26A040' was replaced with 'TMP1 Flow'.

The third step at the beginning of a new project requires importing the dataset into the multivariate data analysis software. From here all variables have to be centered and scaled to unit variance. This step is very important since variables often have substantially different ranges. For example, the paper machine speed may range from 4200 to 4500 ft/min, while the white water consistency may only vary from 0 to 0.50%. Centering and scaling to unit variance ensure that all variables are equally and appropriately weighted before starting.

The fourth and final step before commencing the modelling is to perform a quick review of the time series plot of each variable. This allows for further data cleansing by identifying for instance a time period where a control valve may not have been functioning properly. The observations affected by such a problem can quickly be removed so that they do not incorrectly impact the analysis.

With the dataset prepared as described above PCA and PLS modelling were used to analyze the data. PCA is a technique used to summarize many variables, which are usually correlated, into a few latent variables. PCA does not differentiate between process and response variables, rather it takes all the variables in the data set as one matrix 'X'. PLS modelling combines different mathematical techniques to relate two data sets, process variables 'X' and response variables 'Y'. With PLS modelling it is possible to identify which process variables are responsible for the variability in a process response such as paper quality or retention aid use.

The two most valuable plots in both PCA and PLS modelling are the score plot and loading plot. The score plot is a summary of the relationships among observations. The loading plot is a means to interpret the patterns seen in a score plot. The two plots are complementary and superimposable, a direction in one plot corresponds to the same direction in the other [1].

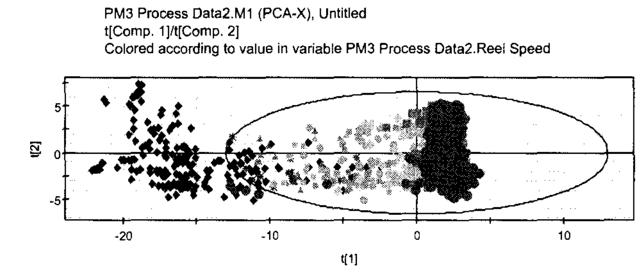

Figure 1. PCA score plot of the initial data set showing periods of paper machine downtime in the outliers to the left. Each point represents one observation time.

The score scatter plot (Fig. 1) of the first PCA shows a group of extreme outliers to the left. A contribution plot of one of these observation points (Fig. 2) shows that this is a period where the reel break status flag was on and the speed was low or at zero signifying a paper break or machine shutdown. All of these data points were removed to ensure the data observation points only included times where the paper machine was actually operating. A score scatter plot with these observation points removed is shown in Figure 3.

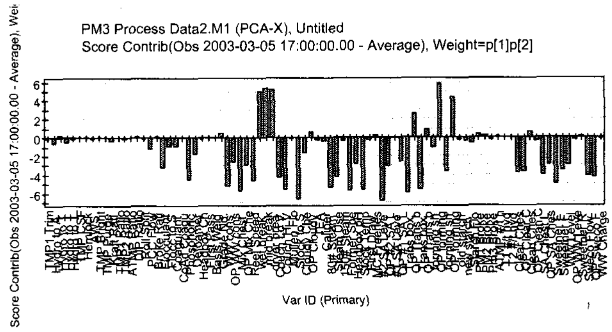

Figure 2. Contribution plot for an extreme outlier in the PCA score plot. This plot highlights the differences, in scaled units, for all the model terms, between the selected outlier and the average observation. The plot above indicates that paper machine break status and low reel speed were the contributing factors.

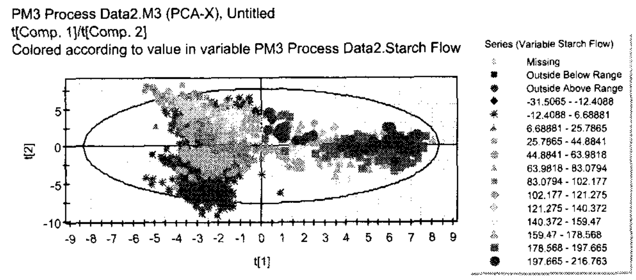

Figure 3. PCA score plot with paper machine downtime outliers removed. This provides a map of how the different observation points are related to each other.

With the dataset cleaned and all major outliers removed a number of PLS models were developed and are discussed in the Results & Discussion section of this paper.

RESULTS AND DISCUSSION

A Principle Components Analysis and three PLS models were developed and will be discussed in this section. The first two PLS models describe the variation in one Y output variable, the first uses the main flocculant addition as the Y variable and the second uses coagulant as the Y variable. It was found that coagulant and flocculant addition were very highly correlated so a final PLS model was developed using all 4 retention aid flows as Y variables.

The actual models developed had over 80 process variables and 3000 observation points, however, for ease of discussion and graphical clarity, the models discussed in this paper have been reduced down to 35-40 variables.

Principle Components Analysis

The 33 components of the PCA explain 87.4% of the variation in the data.

Score scatter plots can reveal distinctly different operating conditions and drifts in the data. For example, in Figure 3, there are distinctive clusters in the graph indicating different operating conditions. Of interest is that the cluster in the far right contains more recent observations. By colouring each data point by ranges of different variables we can learn that this cluster is a region where starch application is high, 150# steam use is down, the white water consistency set-point is low, defoamer use has been reduced, and the machines is running well causing the clean white water chest to overflow.

PLS Model on Flocculant Addition

A 5-component PLS model, M6, was developed were the X input variables were able to explain 71% of the variation of the flocculant addition rate (Y variable). The Model Overview Plot (Fig. 4) displays R^2Y(cum) and Q^2(cum). The R^2Y(cum) is the fraction of the variation of Y(flocculant) that is explained by the model after each component. The Q^2(cum) is the fraction of the variation of Y that can be predicted by the model according to the cross-validation. Values of both of these measures close to 1.0 indicate an excellent model [1].

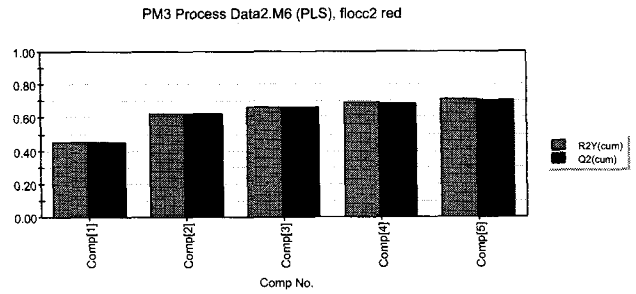

Figure 4. Model overview plot for M6. The R2Y of the five component model is 0.71 and the Q2 is 0.701.

A score scatter plot of the model (Fig. 5) shows good hotelling of the data. This means that the majority of the data falls within the hotelling ellipse drawn on a score scatter plot which represents the 95% confidence level for observations belonging to the model. The loading scatter plot (Fig. 6) shows that increased flocculant flow is highly correlated with high TMP1 use, increased PCC use and coagulant addition (as seen by their location in the top right quadrant of the loading plot). On the other hand, increasing DIP and Kraft use have the opposite effect and tend to result in a decease in flocculant use. When variables are negatively correlated they are positioned on opposite sides of the plot origin, in diagonally opposed quadrants. Variables that fall near the origin of the loading plot have little to no impact on the flocculant addition. It is important to point out that observation points that fall close together on a score scatter plot have similar properties while variables that fall close together on a loading scatter plot are also correlated.

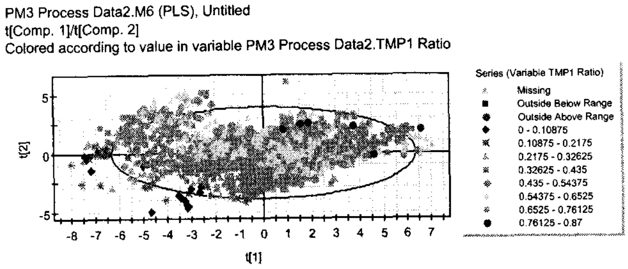

Figure 5. Score scatter plot for M6 with flocculant addition as the Y variable. The graph has been coloured to show periods of increased TMP1 use trending from the left to right.

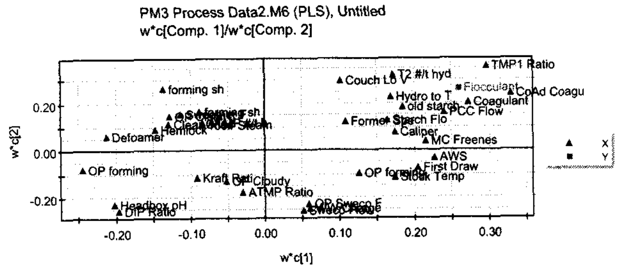

Figure 6. Loading scatter plot for M6 with flocculant addition as the Y variable. The cluster of flocculant, coagulant, PCC and TMP1 Ratio indicate that these variables are highly correlated.

The variable importance plot (Fig. 7) shows the order of importance for each X variable affecting the flocculant flow. The VIP values reflect the importance of terms in the model with respect to their correlation to the responses (Y) and variables (X).

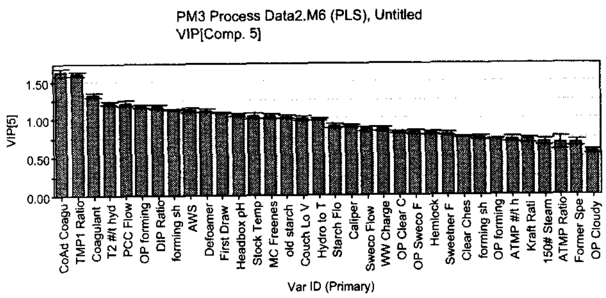

Figure 7. Variable importance plot for M6. CoAd Coagulant is a set ratio to the flocculant addition and is actually a response of high flocculant use. Following this in importance are TMP1 flow, coagulant addition, bleach application and PCC flow.

The coefficients plot for model M6 (Fig. 8) is a useful graph to observe the directional impact of each variable, X, on the flocculant flow, Y. A positively correlated variable is shown by a positive coefficient and a negatively correlated variable is shown by a negative coefficient.

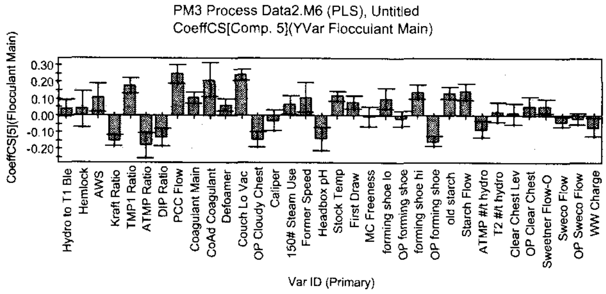

Figure 8. Coefficients plot for M6. This graphs shows that higher PCC use and lower DIP use will result in increased flocculant use.

PLS Model on Coagulant Addition

A 4-component PLS model, M8, was created describing 54.4% of the variability in coagulant addition. The score scatter plot (Fig. 9) and loading scatter plot (Fig. 10) for M8 show a positive correlation of coagulant use with increasing TMP1 application, PCC use and other retention aid flows. From the loading scatter plot the impact of Alaskan White Spruce (AWS) and Hemlock as portions of the TMP furnish can also be seen. AWS was only added to the TMP at rates from 0 to 20% of the overall chip furnish however, with its significantly higher extractives content it had a very large impact on the retention aid use on the paper machines.

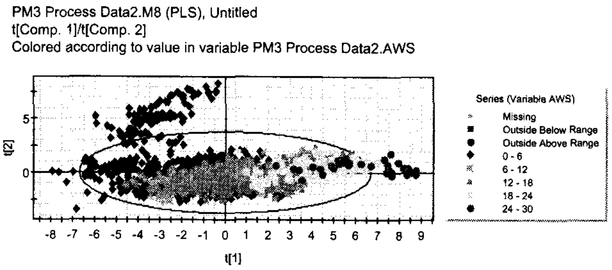

Figure 9. Score scatter plot for M8 with main coagulant addition as Y variable. A period of low TMP water purge caused the cluster of observations in the upper left quadrant. These data points, although outliers, were left in the analysis because of important information they may contain about the impact of TMP purge rate.

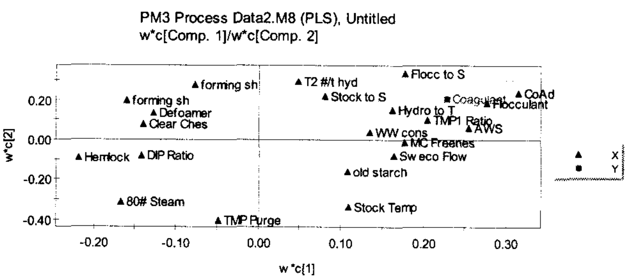

Figure 10. Loading scatter plot for M8. The location of AWS in the top right quadrant near the coagulant flow indicates that there is a strong positive relationship between the two variables.

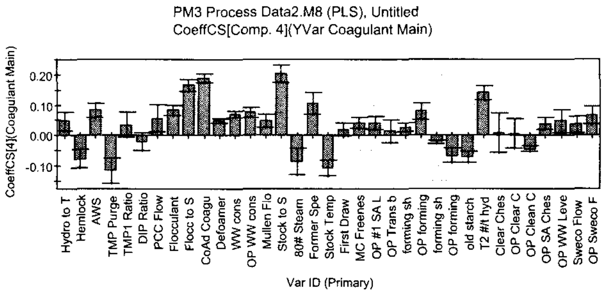

Figure 11. Coefficients plot for M8. This graph clearly shows the positive relationship of Alaskan White Spruce and the negative relationship of Hemlock on the coagulant use.

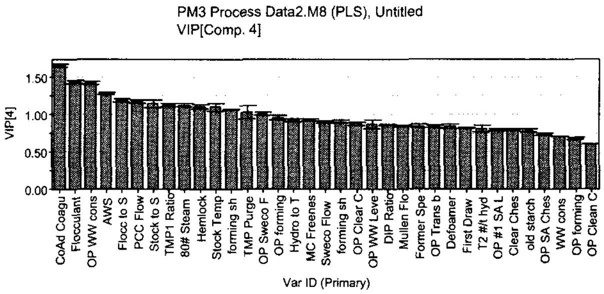

Figure 12. Variable importance plot for M8 showing that after the other retention aid chemicals, AWS, PCC and TMP1 had the largest impact on coagulant use.

PLS Model on all Retention Aid Additions

A final 9-component PLS model explaining 57.7% of the variation in the four retention aid flows was developed. This model highlights the most significant process variables affecting the retention aid use. These were TMP1 Flow, TMP Furnish Species and PCC use. The variable importance and coefficients plots are shown below in Figures 13 and 14 respectively.

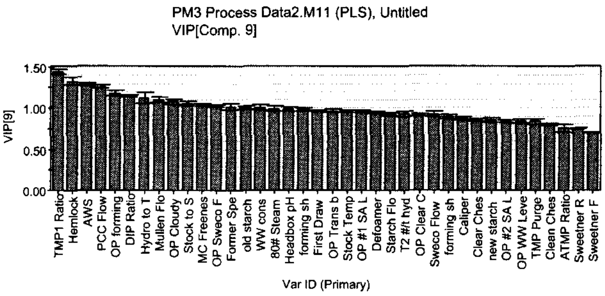

Figure 13. Variable Importance Plot for M11 highlighting the importance of TMP1 Ratio, TMP chip species, PCC flow and DIP application on PM3.

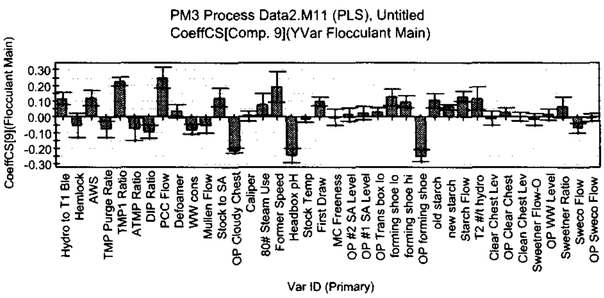

Figure 14. Coefficients Plot for M11. This graph shows the positive relationship between AWS, TMP1, Hydro bleach and PCC flow with retention aid use.

CONCLUSIONS

Utilizing multivariate data analysis tools has been a great benefit to NorskeCanada Crofton Division by providing a better scientific understanding of our wet end processes. Since initial analyses were done, #3 Paper Machine has started base loading with de-inked pulp. This has helped to stabilize the Paper Machine and maintain a lower white water consistency set point, improving formation and linting propensity. In addition to stabilizing the amount of DIP used, PCC usage has been reduced and Alaskan White Spruce has been removed from the TMP furnish. The combination of these changes has allowed for a reduction in retention aid use on #3 Paper Machine as predicted by the MVDA models.

These positive results which confirm the initial model predictions and have resulted in cost savings for the paper machine have given the mill confidence in using multivariate data analysis software to accurately troubleshoot process problems.

The next step to using the full potential of the models discussed in this paper is to verify their ability to predict process changes and to use them as on-line tools to monitor real-time process changes.

REFERENCES

1. ERIKSSON, L. et al., "Multi- and Megavariate Data Analysis", [illegible] (2001).